A SELECTIONIST APPROACH TO REINFORCEMENT

JOHN W. DONAHOE, JOSE E. BURGOS, AND DAVID C. PALMER

UNIVERSITY OF MASSACHUSETTS AT AMHERST AND SMITH COLLEGE

We describe a principle of reinforcement that draws upon experimental analyses of both behavior and the neurosciences. Some of the implications of this principle for the interpretation of behavior are explored using computer simulations of adaptive neural networks. The simulations indicate that a single reinforcement principle, implemented in a biologically plausible neural network, is competent to produce as its cumulative product networks that can mediate a substantial number of the phenomena generated by respondent and operant contingencies. These include acquisition, extinction, reacquisition, conditioned reinforcement, and stimulus-control phenomena such as blocking and stimulus discrimination. The characteristics of the environment–behavior relations selected by the action of reinforcement on the connectivity of the network are consistent with behavior-analytic formulations: Operants are not elicited but, instead, the network *permits* them to be guided by the environment. Moreover, the guidance of behavior is context dependent, with the pathways activated by a stimulus determined in part by what other stimuli are acting on the network at that moment. In keeping with a selectionist approach to complexity, the cumulative effects of relatively simple reinforcement processes give promise of simulating the complex behavior of living organisms when acting upon adaptive neural networks.
Key words: reinforcement, selectionism, neural networks, evolution, interpretation

Within evolutionary biology, Darwin's great insight was that the complexity of species arose as the cumulative product of the repeated action of relatively simple biobehavioral processes, most notably those whose effects are functionally described by the principle of natural selection. More broadly conceived, Darwin's account was the first comprehensive proposal whereby higher level complexity could be interpreted as the "unintended" effect of the repeated action of lower level processes. Complexity was unintended in the sense that no external agencies or order-imposing principles were needed to oversee its production. Instead, complexity emerged as a by-product of the three-step sequence of *variation* of behavioral and morphological characteristics, *selection* by the environment of those characteristics that affected reproductive fitness, and *retention* of the selected variations via the mechanisms of heredity (cf. Campbell, 1974; Mayr, 1982). These retained characteristics

were then available to contribute to the variation upon which subsequent selections acted, with complexity as a possible cumulative outcome.

Darwin's approach to complexity—*selectionism*—has since been explicitly pursued throughout biology and is implicit in accounts of complex phenomena in other natural sciences (cf. Campbell, 1974). For example, the formation of planetary systems as the result of gravitational and other processes acting on a swirling cloud of interstellar dust particles exemplifies selectionism (cf. Gehrz, Black, & Solomon, 1984). For a planet to orbit the sun, it must have just enough velocity tangent to its orbit to compensate for the tendency to fall toward the sun. If its velocity is too great, it escapes from orbit; if too low, it spirals into the sun. The planets achieve the velocities required to maintain their orbits, and those bodies with the requisite velocities are all that remain to be observed. Gravitational force produces organized complexity as a by-product, with most of the primordial matter collapsing into the sun and planets or escaping the solar system altogether.

Note, however, that although gravitation, together with other physical processes, may be *sufficient* to account for the formation of planets, the specific arrangement of planets that characterizes our solar system is not a *necessary* consequence of their action. The same processes—acting on different initial conditions and, hence, in different sequences—are com-

petent to produce planetary systems of many different configurations. Moreover, the order that we now observe may not be stable, as, over time, matter assumes new orbits about the body that it orbits or reveals itself to be moving chaotically. Whether concerned with planetary systems or species, the cumulative product of selection processes may be not only complex but also diverse (cf. Donahoe & Palmer, 1989). The variation among species is particularly eloquent testimony to the diversity as well as the complexity of which selection processes are capable. (See Palmer & Donahoe, 1992, for a discussion of other characteristics of the products of selection.)

TOWARD A SELECTIONIST ACCOUNT OF BEHAVIORAL COMPLEXITY

From a selectionist perspective, a principle of reinforcement is central to an account of behavioral complexity. A well-formulated principle of reinforcement should bear the same relation to the emergence of complex behavior as the principle of natural selection bears to the emergence of complex morphology. That is, a principle of reinforcement should prove to be as fundamental and as fruitful to understanding the origins of complex behavior in individual organisms as the principle of natural selection has proved to be in understanding the origins of complex morphologies in species (Donahoe, Crowley, Millard, & Stickney, 1982).

On the behavioral level, experimental analysis leading to a principle of reinforcement seeks to identify the conditions under which the behavior of the individual organism comes to be guided by the environment. Indeed, experimental analysis has identified such conditions—the brief temporal intervals between the environmental, behavioral, and reinforcing events that define three-term contingencies (e.g., Ferster & Skinner, 1957; Skinner, 1938) and the evocation by the reinforcing stimulus of behavior that would not otherwise occur in that environment (Kamin, 1968, 1969; cf. Rescorla, 1969; Rescorla & Wagner, 1972). When appropriate temporal relations occur between these events in proximity to a reinforcer-induced behavioral change (i.e., a behavioral discrepancy), the environmental guidance of be-

havior is modified (Donahoe et al., 1982; Stickney & Donahoe, 1983).

Historical Reception of a Selectionist Approach

Environmental conditions have been identified under which selection by reinforcement occurs and plausible interpretations based on those findings have been provided for a wide range of complex environment–behavior relations. Nevertheless, few outside the behavior-analytic community accept the proposition that complex behavior can be understood as the cumulative product of relatively simple reinforcement processes. Why has selection by reinforcement not been widely accepted as the best extant account of behavioral complexity, whereas natural selection has triumphed in its domain? The answer to this question is potentially important because an appreciation of the scope of a principle of reinforcement may depend upon the existence of circumstances analogous to those that preceded the acceptance of the principle of natural selection. What were those circumstances?

Darwin (and, independently, Alfred Wallace) proposed the principle of natural selection in 1859 as the central insight into what he called "that mystery of mysteries," the origin of species. What is insufficiently appreciated is that, although the notion of evolution was generally accepted by the scientific community, natural selection as the process whereby evolution occurred was not embraced until the 1930s (e.g., Dobzhansky, 1937)—some 70 years later! This period of "the eclipse of Darwinism" has been discussed elsewhere (e.g., Bowler, 1983; Hull, 1973) and has been commented upon in this journal (Catania, 1987). Although a number of circumstances contributed to the acceptance of Darwinism, two are especially important. First was the rediscovery of Mendel's work, which led to the identification of the biological mechanisms, the genetic bases of heredity, through which Darwin's functional account could be realized. Second was the development of population genetics, with its more formal techniques—statistics as developed by Ronald Fischer, J. B. S. Haldane, and Sewell Wright and, much later, computer simulation (e.g., Maynard Smith, 1982)—for tracing the course of selection. These more rigorous techniques provided a means for exploring the implications of nat-

ural selection that were more compelling than Darwin's verbal interpretations. The integration of the mechanisms of heredity with population genetics provided a persuasive account of evolution through natural selection and formed what is now known as the "modern synthesis" or the "synthetic theory" of evolution.

What lessons bearing on the acceptance of a principle of selection by reinforcement may be drawn from the record of the acceptance of natural selection? If a historical parallel holds, then the acceptance of a principle of selection by reinforcement awaits the identification of its biological mechanisms and the development of techniques for interpreting its implications that are more rigorous than verbal interpretation (cf. Donahoe & Palmer, 1989). This is not to say that either the identification of biological mechanisms or the development of more formal interpretative techniques are logically necessary in order for selection by reinforcement to be preferred to alternative accounts of behavioral complexity. Rather, the historical record suggests that both may be "psychologically" necessary for the general acceptance of reinforcement as the key insight into the origins of behavioral complexity. Accordingly, the experimental analysis of behavior should be supplemented (not replaced) by experimental analyses of the neurosciences, and the resulting synthesis should be interpreted using more formal techniques. The integration of behavioral and neuroscientific findings would constitute a new modern synthesis that might claim behavioral complexity as its domain just as the synthetic theory of evolution now claims morphological complexity. We refer to the synthesis of behavior analysis and the neurosciences as the *bio*behavioral approach (see Figure 1).

Some may regard a call for the integration of behavior analysis with the neurosciences as questioning one of Skinner's major accomplishments, the establishment of an independent science of behavior. This would be a misperception of our position and, more importantly, of Skinner's. As has been noted elsewhere (Donahoe & Palmer, 1989), to argue for an integration of the experimental analysis of behavior and physiology in no way undermines the independence of behavior analysis. Behavior analysis remains as independent of physiology as physiology is of biochemistry, and as interdependent as well.

Although Skinner was educated at least as much in the biology as the psychology of the time, he engaged in almost no experimental work at the physiological level. As someone committed to the study of behavior, he regarded the coordination of behavior and physiology as dependent on the prior establishment of a science of behavior (Skinner, 1938, pp. 423–424). Also, Skinner had little more confidence in the science of physiology available when he began his work than he did in the extant science of behavior. He viewed much of physiology and almost all of psychology as concerned with what he called the "Conceptual Nervous System" (Skinner, 1938, pp. 421), that is, a "nervous system" whose structures and processes were inferred from observations at other levels of analysis, not the real nervous system that could be directly subjected to experimental analysis. In short, Skinner's lack of concern for things physiological was grounded in strategic and pragmatic considerations, not in principled reservations about the potential relevance of physiology to behavior, and, equally important, of behavior to physiology.

TOWARD A *BIO*BEHAVIORAL PRINCIPLE OF REINFORCEMENT

In keeping with the foregoing overview of the circumstances that preceded the acceptance of the principle of natural selection, the twin goals of the present paper are to formulate a principle of reinforcement informed by both behavioral and neuroscientific research and, then, to explore some of its implications using a more formal interpretive technique than verbal interpretation (i.e., computer simulation via adaptive neural networks).

A Unified Principle of Reinforcement

As already noted, the experimental analysis of behavior has identified two sets of conditions that are required for selection by reinforcement: (a) brief temporal intervals between the environmental, behavioral, and reinforcing events of the three-term contingency and (b) a reinforcing stimulus that evokes a behavioral change or discrepancy. On this view, the sensitivity of organisms to relations defined over
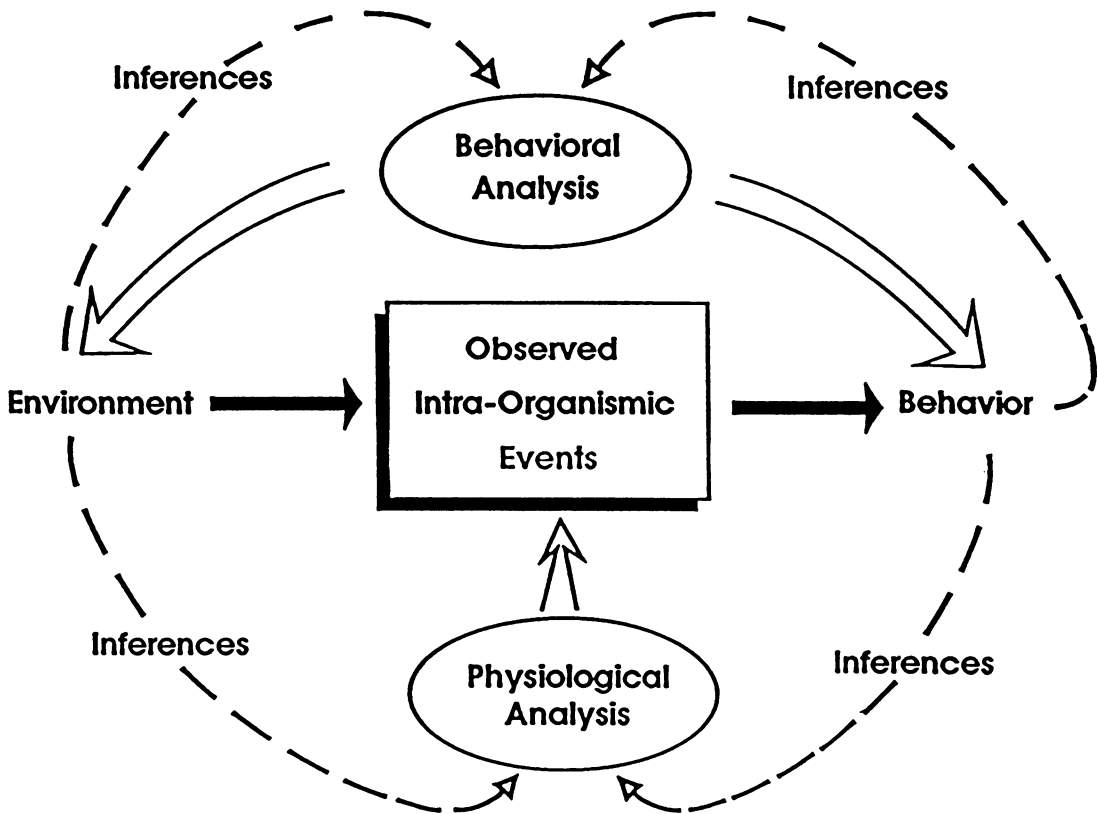
Biobehavioral Approach

Fig. 1.   A biobehavioral approach to the analysis and interpretation of complex behavior. The experimental analysis of behavior, which is concerned with the effects of environmental manipulations on behavior, is supplemented by the experimental analysis of physiology, which is concerned with the effects of intraorganismic manipulations on intraorganismic events and behavior.

longer time intervals (so-called molar relations or correlations) is the *cumulative* effect of moment-to-moment relations among environmental, behavioral, and reinforcing events (Donahoe et al., 1982; Rescorla & Wagner, 1972). That is, sensitivity to correlation is the *emergent* product of sensitivity to contiguity.

This molecular view is consistent with Skinner's approach to selection by reinforcement. For example, in *Schedules of Reinforcement* (Ferster & Skinner, 1957; cf. Skinner, 1981), it states

> A more general analysis is . . . possible which answers the question of *why* (emphasis in original) a given schedule generates a given performance. . . . It does this by a closer analysis of the actual contingencies of reinforcement prevailing under any given schedule. . . . The

only contact between [the schedule] and the organism occurs at the *moment* (emphasis added) of reinforcement. . . . Under a given schedule of reinforcement, it can be shown that at the moment of reinforcement a given set of stimuli will usually prevail. A schedule is simply a convenient way of arranging this. (pp. 2–3)

The phrase "moment of reinforcement" appears at several other points in the introduction to the study of reinforcement schedules, and many examples of moment-to-moment analyses appear in discussions of the behavioral effects of various schedules (see also Morse, 1966; Skinner, 1938, 1948).

Again, none of this denies the existence of molar regularities (e.g., between mean response rate and mean reinforcement rate, as in the matching law; Herrnstein, 1970). Nor

should the emphasis upon moment-to-moment contingencies be seen as reducing the contribution that an appreciation of molar regularities makes to the verbal interpretation of behavior (e.g., the implication of the matching law that dysfunctional behavior can be reduced by reinforcing alternative responses rather than punishing the dysfunctional behavior). Instead, as already noted, the present view affirms that these molar regularities are the cumulative and emergent products of moment-to-moment relations among events defined over brief time intervals. Consistent with Skinner's advocacy of an analysis of molecular contingencies, recent theoretical work has progressively moved toward moment-to-moment interpretations of previously uncovered molar regularities (e.g., Herrnstein, 1982; Heth, 1992; Hinson & Staddon, 1983; Shimp, 1969; Silberberg, Hamilton, Ziriax, & Casey, 1978; Staddon & Hinson, 1983; Vaughan, 1981). Although future work must determine whether molecular accounts can encompass all molar regularities (cf. Nevin, 1979; B. Williams, 1990), current research suggests that many molar regularities fall within the reach of moment-to-moment analyses.

Operant–respondent distinction. Paradoxically, the consistent application of a moment-to-moment analysis undermines the cogency of the distinction between operant and respondent conditioning as fundamentally different *types* of conditioning requiring different principles for their understanding (cf. Skinner, 1935b, 1937). Consider the prototypical respondent conditioning procedure of Pavlov in which the ticking of a metronome was paired with the introduction of meat powder into the mouth of a dog. Although the reinforcing stimulus (here, meat powder) occurred independently of behavior, it is inescapably true that some behavior must necessarily have preceded the reinforcing stimulus. For example, the dog's ears might prick up or its head turn toward the sound of the metronome immediately before receiving the meat powder. As Schoenfeld has emphasized, reinforcers are necessarily introduced into an ongoing "stream" of behavior (e.g., Schoenfeld et al., 1972). Thus, although Pavlov's dogs need not have behaved in any particular manner prior to receiving the reinforcing stimulus, they were nevertheless behaving in some manner, even if they were standing perfectly still. Responses as well as

stimuli necessarily precede reinforcers in a respondent procedure although, over time, only stimuli reliably precede the reinforcer.

Similarly, in the prototypical operant procedure of Skinner, lever pressing preceded food but some stimulus must necessarily have been sensed immediately prior to the reinforcer. (For an insightful examination of Skinner's abandonment of the stimulus in his treatment of operant conditioning, see Coleman, 1981, 1984.) For example, just prior to lever pressing the rat might have seen the lever, or smelled some odor within the chamber, or sighted the houselight while attempting to climb out of the chamber and, in so doing, "inadvertently" emitted the criterion response. Although the rat need not have sensed any particular stimulus prior to pressing the lever and receiving the food, *some* stimulus must have been sensed immediately prior to the response and the reinforcer. Thus stimulus events necessarily precede the reinforcer in the operant procedure although, over time, only responses *reliably* precede the reinforcer. (Of course, specific stimuli may be scheduled to precede the reinforcer in a discriminated operant procedure.)

What is crucial to note in the foregoing description of the procedures that implement respondent and operant contingencies is that at the moment when food occurs (i.e., at the moment of reinforcement), both procedures necessarily contain a sequence of the same types of events—stimulus, response, and reinforcer. In both procedures some stimulus must have been sensed and some response must have occurred prior to the reinforcer (see Figure 2). Although the experimenter manipulates different stimuli and measures different responses in respondent and operant procedures, the learner is exposed to a similar sequence of events in either case (Donahoe et al., 1982). If comparable momentary sequences of events occur in the two procedures and if selection by reinforcement is governed by moment-to-moment relations between events, then the basis upon which a principle of reinforcement may differentiate between the procedures is eliminated. The two procedures cannot require different "laws of learning" because, even if different laws existed, no basis would exist at the moment of selection with which the organism could "decide" which set of laws to invoke. This does not mean that the two procedures may not have very different *cumulative* effects

ENVIRONMENT S_1 S_2 S_3 • • • S_i • • • S_m • • •

CLASSICAL

US — UR

OPERANT

BEHAVIOR R_1 R_2 R_3 • • • R_j • • • R_n • • •
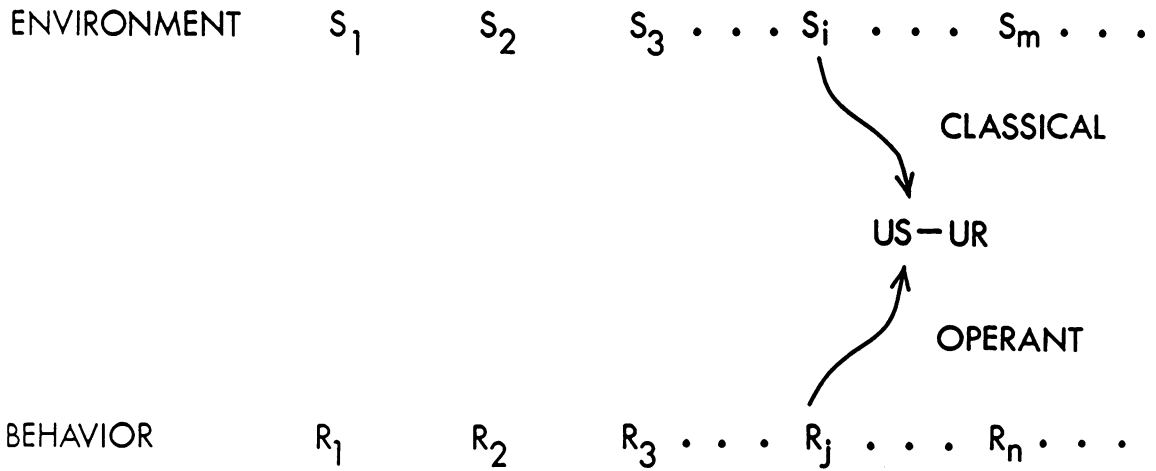
Fig. 2.   An organism is immersed in a stream of environmental events, or stimuli (S), in whose presence the organism is continuously behaving, or responding (R). With a respondent, or classical, contingency the occurrence of an eliciting, or unconditioned, stimulus (US) is contingent on an environmental event, but some behavioral event must necessarily precede the US. The US functions as the putative reinforcing stimulus. With an operant, or instrumental, contingency the occurrence of the US is contingent on a behavioral event, but some environmental event must necessarily precede the US. The US functions as the putative reinforcing stimulus. With an operant, or instrumental, contingency the occurrence of the US is contingent on a behavioral event, but some environmental event must necessarily precede the US. (The two types of contingencies are indicated by wavy lines with arrows.) In the operant procedure, the responses that are candidates for control by the environment include both the operant (R) and the elicited, or unconditioned, response (UR).

on behavior because of the differences in which events reliably precede the reinforcer in the two procedures. However, it does mean that a commitment to a moment-to-moment analysis requires the formulation of a common principle of reinforcement that is competent to produce the different behavioral outcomes of respondent and operant contingencies as its emergent product.

To repeat, a commitment to a moment-to-moment analysis does not imply that the cumulative effect of selection by reinforcement is the same for the two contingencies. Quite the contrary, as Skinner consistently pointed out; the net effects of operant and respondent contingencies differ in ways that have profound implications for the emergence of complex behavior, and we concur with that view. Most importantly, responses from the full behavioral repertoire of the organism are candidates for selection by reinforcement with an operant contingency. By contrast, with a respondent contingency the eligible responses are confined to those that are already elicited by specific environmental stimuli.

A single reinforcement principle—the unified reinforcement principle—has been proposed to accommodate conditioning with both respondent and operant contingencies while also yielding, as its cumulative product, the emergence of the different behavioral outcomes that typify the two contingencies (Donahoe et al., 1982). The unified reinforcement principle holds that whenever a behavioral discrepancy occurs, other things being equal, all those stimuli preceding the discrepancy will acquire control over all those responses occurring immediately prior to and contemporaneous with the discrepancy. In a respondent contingency, the most reliably occurring stimulus is the conditioned stimulus (e.g., the sound of the metronome) and the most reliably occurring responses are orienting responses (e.g., turning toward the source of the tone) and unconditioned responses (e.g., salivation). In a discriminated operant contingency, the most reliably occurring stimulus is the discriminative stimulus (e.g., the presence of a tone only during periods when lever pressing produces food) and the most reliably occurring response is, in addition to the orienting response to the discriminative stimulus and the unconditioned response to the reinforcing stimulus, the operant itself (Donahoe et al., 1982; see also Mackintosh, 1983; Staddon, 1983). The net outcome of either contingency depends on the interac-

tions, if any, among the various responses and whatever constraints arise from natural selection (so-called "biological constraints on learning"). The unified reinforcement principle is based upon the experimental analysis of behavior and provides the conceptual framework that we shall supplement with findings from the experimental analysis of the neurosciences.

Behavioral Constraints on a Principle of Reinforcement

In addition to the requirement that a single principle of reinforcement accommodate conditioning with both respondent and operant contingencies, behavioral considerations impose several additional constraints upon a biobehavioral account of reinforcement. First, a selection principle must, in general, permit the stimuli and responses that enter into selected environment–behavior relations to include as candidates a wide range of the stimuli and responses preceding the reinforcer. Without maintaining the eligibility for selection of a wide range of stimuli and responses, relatively arbitrary environment–behavior relations could not be acquired, with a corresponding limitation on the potential complexity of relations that could be selected. Although maintaining a potentially large pool of candidate stimuli and responses facilitates the selection of complex environment–behavior relations, the effect of this constraint also allows the acquisition of superstitions of both the first (Skinner, 1948) and second kinds (Morse & Skinner, 1957). That is, the selected relation may include response topographies containing responses that are not necessary to produce the reinforcer and stimulus complexes containing stimuli that do not necessarily precede the reinforcer. Selection by reinforcement no more ensures the isolation of the necessary and sufficient conditions for reinforcement in any given situation than the principle of natural selection ensures the reproductive fitness of any given individual. It is not in the nature of selection processes to guarantee the parsing of events into necessary and sufficient relations.

A second and countervailing behavioral constraint is that the cumulative effect of selection by reinforcement must, more often than not, converge on stimulus and response classes whose members are most reliably correlated with the reinforcer. That is, in general, the cumulative effect of selection must yield a valid theory of the world in which the organism lives. If this constraint were not satisfied, natural selection could not have favored the neural mechanisms mediating selection by reinforcement. Thus a principle of reinforcement may not be immune to superstitions but, at the same time, it must, more often than not, have the cumulative effect of veridically extracting the correlations between environmental and behavioral events that precede reinforcers (cf. Stone, 1986). Note that, given these boundary constraints, the product of selection by reinforcement is a relation between *classes* of stimuli and *classes* of responses, as Skinner had anticipated (Skinner, 1935a). Moreover, these classes are "fuzzy" classes, in that no one member of either class may be a necessary component of the selected environment–behavior relation. (See Palmer & Donahoe, 1992, for a discussion of the far-reaching implications of Skinner's insight regarding reinforcement as the selection of relations between classes.)

Acquired reinforcement. The final behavioral constraint considered here is that a reinforcement principle must accommodate the selecting effect of acquired as well as unconditioned reinforcers. The term *acquired reinforcer* is used in preference to either conditioned (secondary) reinforcer for operant contingencies or higher order conditioning for respondent contingencies because these latter terms imply a fundamental distinction between the theoretical treatments of the two contingencies, a distinction that is not honored by the unified reinforcement principle. Procedurally, respondent and operant conditioning are undeniably distinguishable; conceptually, the same reinforcement principle seeks to encompass both.

From the present perspective, any stimulus that evokes a change in behavior (i.e., a behavioral discrepancy) can potentially function as a reinforcer to select environment–behavior relations. No fundamental distinction is made between stimuli that evoke behavior because of prior selection by the ancestral environment, as with unconditioned stimuli, and stimuli that evoke behavior because of prior selection by the individual environment, as with conditioned and discriminative stimuli. Regardless of the origin of their ability to evoke responses, stimuli function as reinforcers to the extent that their occurrence produces a *change* in ongoing behavior.

The central observation affirming a critical

role for behavioral discrepancy in selection by reinforcement is the phenomenon of blocking. In blocking, a stimulus standing in a favorable temporal relation to a putative reinforcer will not acquire the capacity to function as either a conditioned stimulus (Kamin, 1968, 1969) or a discriminative stimulus (vom Saal & Jenkins, 1970) if that stimulus is paired with the reinforcer in the presence of a second stimulus that already evokes the reinforcer-elicited response. For example, when a light is presented in simultaneous compound with a tone and both stimuli are followed by food, the light will not come to evoke salivation if the tone has previously been paired with food. Because the tone already evokes salivation as a result of prior conditioning, no *change* in behavior occurs when salivation is also evoked by food in the presence of the light–tone compound stimulus. Consequently, selection by the putative reinforcer cannot occur, and the light does not acquire control over salivation.

Because the blocking procedure prevents a stimulus from functioning as either a conditioned or a discriminative stimulus, the blocked stimulus should also be prevented from functioning as an acquired reinforcer. Experimental analysis confirms that when the discriminative function of a stimulus has been blocked, its ability to function as an acquired reinforcer is also blocked, even though the stimulus has been paired with an unconditioned reinforcer many times (Palmer, 1987). Thus both acquired and unconditioned reinforcers must evoke behavioral change if they are to function as reinforcers. This view is consistent with earlier behavior-analytic work indicating that acquired (conditioned) reinforcers must first have the status of discriminative stimuli (e.g., Keller & Schoenfeld, 1950; cf. Dinsmoor, 1950; Thomas & Caronite, 1964). Whether a stimulus that evokes behavior functions as a reinforcer for a particular environment–behavior relation depends, for both acquired and unconditioned reinforcers, upon any interactions between the response evoked by the putative reinforcing stimulus and the operant upon which it is contingent (e.g., Long, 1966; cf. Donahoe et al., 1982).

Neuroscientific Constraints on a Principle of Reinforcement

In addition to behavioral constraints, a biobehavioral principle of reinforcement must also be constrained by relevant neuroscientific findings. A moment-to-moment account of reinforcement at the behavioral level is congenial to supplementation by a neuroscientific account because, whatever their nature, the neuronal changes mediating reinforcement necessarily occur on a moment-to-moment basis. As Skinner (1938) stated, "I agree with Carmichael [1936] that 'those concepts which do not make physiological formulation impossible and which are amenable to growing physiological knowledge are preferable, other things being equal, to those that are not so amenable'" (p. 440).

What is known about the neural systems and cellular processes mediating reinforcement? Although a comprehensive answer to this question is beyond the scope of the present paper and some important questions remain unanswered (cf. Donahoe & Palmer, in press; Krieckhaus, Donahoe, & Morgan, 1992), the major outlines may be given here.

Neural selection of environment–behavior relations. When neurons whose cell bodies are in the ventral tegmental area (VTA, see Figure 3) are electrically stimulated after an operant has occurred, the strength of the operant is increased. Thus VTA stimulation functions as a reinforcer. Moreover, experimental work indicates that VTA neurons are activated by environmental stimuli that commonly function as unconditioned reinforcers, such as the smell and taste of food (see Hoebel, 1988; Trowill, Panksepp, & Gandelman, 1969, for reviews).

As shown schematically in Figure 3, axons from VTA neurons diffusely project throughout the motor association areas of the frontal lobes (Fallon & Loughlin, 1987; Swanson, 1982). The neuromodulator dopamine (DA) is released by VTA fibers, and, because of their widespread projections, dopamine is positioned to affect synaptic efficacies throughout the frontal lobes. The affected synapses are (among others) between presynaptic neurons carrying sensory information from the parietal-temporal-occipital lobes of the cortex that are activated by environmental events (e.g., S_{VISUAL} in Figure 3) and postsynaptic neurons in the frontal lobes that lead ultimately to behavior. Cellular research has shown that the introduction of dopamine into synapses immediately *after* a postsynaptic neuron has been activated by a presynaptic neuron produces long-lasting changes in synaptic efficacies. That

is, the ability of the presynaptic neuron to initiate activity in the postsynaptic neuron is increased (Stein & Belluzzi, 1988, 1989; see also Beninger, 1983; Bliss & Lomo, 1973; Hoebel, 1988; Irikl, Pavlides, Keller & Asanuma, 1989; Kety, 1970; Levy & Desmond, 1985; Wise, 1989; Wise & Bozarth, 1987). Because presynaptic neurons may be activated by sensory input and postsynaptic neurons may lead to behavior, the cumulative effect of the introduction of dopamine into synapses between coactive neurons in the motor association cortex is to select a neural system that mediates environment–behavior relations scheduled for reinforcement.[1]

Because dopamine is released into synapses between many coactive neurons, some of the affected synapses may mediate environment–behavior relations other than those necessary for the reinforcer. However, among the coactive neurons must be those that mediated the reinforced relation, because, without the occurrence of the criterion response and the stimuli present when it was emitted, the reinforcing stimulus that activated the VTA neurons would never have occurred. And, because it is synapses between these neurons into which dopamine most reliably diffuses, these synapses are, cumulatively, most substantially modified. In short, the proposed neural system for reinforcement operates on a moment-to-moment basis but has the cumulative effect of modifying those synaptic efficacies that most reliably mediate reinforced environment–behavior relations. Such a neural system of reinforcement can fall prey both to stimulus and re-
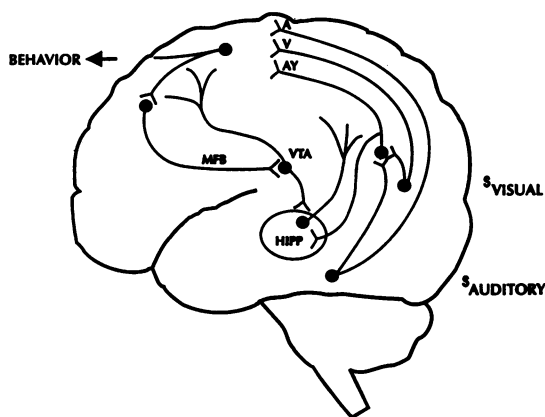

Fig. 3. A schematic diagram of the proposed cortical systems of the brain involved in modifying synaptic efficacies in sensory and motor cortices. Stimuli from various sensory channels (e.g., S_{VISUAL} and $S_{AUDITORY}$) activate neurons in the primary sensory cortex. Axons from the primary sensory cortex ultimately end in synapses on neurons in the motor association cortex (e.g., A and V) and on polysensory neurons in the sensory association cortex. Axons from polysensory cells ultimately end in synapses on neurons in the motor association cortex (e.g., AV) and also on cells within the hippocampus (HIPP). Hippocampal CA1 neurons project diffusely via multisynaptic pathways to the sensory association cortex. Within the motor association cortex, the axons of some neurons ultimately activate effectors that lead to behavior. Axons from some neurons within the motor association cortex project back to the ventral tegmental area (VTA) by means of the medial forebrain bundle (MFB). VTA neurons are activated by unconditioned reinforcing stimuli and project diffusely to the motor association cortex and to the synapses of CA1 hippocampal neurons. (See the text for the functioning of these neural systems in the selection of pathways mediating environment–environment and environment–behavior relations.)

[1] The discussion has focused on the strengthening of connections in the cerebral cortex. In the cortex and some subcortical areas (such as the nucleus accumbens), the levels of dopamine are phasically increased when reinforcers occur. In other subcortical areas, and in the evolutionarily older regions of the brain generally, dopamine appears to be tonically present. Most neuroscientific analyses of conditioning in mammals have been conducted with preparations in which simple conditioning does not require an intact cortex (e.g., nictitating membrane conditioning in the rabbit; Thompson, 1990). Accordingly, the contribution of dopamine to changes in synaptic efficacies is less obvious because levels are more stable (cf. Lowel & Singer, 1992; Singer & Rauschecker, 1982). However, when the action of dopamine is blocked by appropriate receptor antagonists in these preparations, its effects are consistent with the proposal presented here (e.g., Gimpl, Gormezano, & Harvey, 1979; Marshall-Goodell & Gormezano, 1991). That is, the acquisition of conditioned responses is impaired.

sponse superstitions (as when irrelevant sensory inputs or response outputs occur prior to the reinforcer), but it cumulatively selects those environment–behavior relations that most reliably precede reinforcers.

Acquired reinforcement. Together, the neural systems and cellular mechanisms just described provide an account whereby unconditioned stimuli that activate VTA neurons function as reinforcers to select environment–behavior relations. However, such a neural system cannot, by itself, accommodate acquired reinforcers. Acquired reinforcers vary with the particulars of the organism's selection history and, therefore, natural selection cannot provide pathways to the VTA that are "appropriate" for these diverse and changing histories. Natural selection can provide pathways

that have the potential, when acted upon by selection by unconditioned reinforcers, to strengthen differentially the synaptic efficacies along pathways that implement acquired reinforcement. As shown in Figure 3, some of the neurons in the frontal association cortex have axons that project back to the VTA via the medial forebrain bundle (MFB) (Shizgal, Bielajew, & Rompre, 1988; Yeomans, 1988, 1989). If synaptic efficacies to these "feedback" neurons are increased by the action of unconditioned (or previously acquired) reinforcers, then discriminative stimuli become able to function as acquired reinforcers. These stimuli function as acquired reinforcers because they activate VTA neurons through feedback from pathways arising from the frontal lobes rather than through the phylogenetically selected pathways used by unconditioned reinforcers.

As a result of the foregoing process, stimuli that are constituents of previously selected environment–behavior relations lead not only to the emission of behavior but also to the engagement of the neural mechanisms of acquired reinforcement. In this manner, acquired reinforcement facilitates the acquisition of new environment–behavior relations. For example, if backward chaining is used to establish a component of a long behavioral sequence, the stimuli in the first component activate the acquired reinforcement system and thereby function as reinforcers for responses in the second component. The proposed neural system for acquired reinforcement is consistent with findings from the experimental analysis of behavior: Stimuli that function as acquired reinforcers also function as discriminative stimuli.

Neural selection of environment–environment relations. The account of the neural mechanisms of unconditioned and acquired reinforcement of environment–behavior relations assumes that activity arising from sensory regions of the brain is sufficient to distinguish environments in which a given activity is reinforced from those in which it is not reinforced (or differently reinforced). For reasons more fully described elsewhere (Donahoe & Palmer, in press), this may not always be the case. For example, consider a rat for which lever pressing is reinforced during the co-occurrence of an auditory and a visual stimulus, but is not reinforced when either stimulus occurs alone

or when both are absent. Such situations define stimulus patterning or configural conditioning in behavioral research (e.g., Woodbury, 1943; cf. Kehoe, 1988) and exclusive-OR problems (the simplest nonlinearly separable problem) in artificial intelligence research (Rumelhart, Hinton, & Williams, 1986). Some neural mechanisms must exist whereby moment-to-moment processes may cumulatively allow the organism to be sensitive to the correlations among environmental events (e.g., between lights and tones) as well as between environmental and behavioral events. That is, both environment–behavior relations and the environment–environment relations upon which some environment–behavior relations depend must be selected. What neural systems implement the selection of environment–environment relations?

As shown schematically in Figure 3, neuroanatomical studies indicate that axons from neurons in the sensory association cortex, in addition to innervating motor areas of the brain, initiate activity in pathways that provide inputs to the hippocampus. Upon entering the hippocampus, activity is initiated among hippocampal neurons, and ultimately the CA1 hippocampal neurons are activated. Axons from CA1 neurons constitute the major output of the hippocampus and are the origins of multisynaptic pathways that project diffusely back to the sensory association cortex from which the inputs to the hippocampus arose (Amaral, 1987). Because of this arrangement, the output of CA1 neurons is positioned to modulate the functioning of cells throughout the sensory association cortex (Donahoe & Palmer, in press; Krieckhaus et al., 1992).

We propose that diffuse feedback from the hippocampus exerts a neuromodulatory effect on synaptic efficacies in the sensory association cortex that is analogous to the effect of the diffuse VTA-derived reinforcing system on the motor association cortex. That is, diffuse hippocampal feedback strengthens synaptic efficacies between coactive pre- and postsynaptic neurons. If this is the case, then the most reliably affected synapses are those whose activity is correlated with the output of CA1 cells. For example, suppose that an auditory and a visual stimulus occur together, and that their co-occurrence causes a hippocampal output (see Figure 3). The diffusely projected output of

the hippocampus would increase synaptic efficacies to polysensory "audio-visual" cells in the sensory association cortex. As a result, these polysensory cells would become more strongly polysensory (i.e., more reliably activated by the co-occurrence of stimuli from different sensory channels). The cumulative effect of this process is that synaptic efficacies of polysensory cells are modified to reflect the correlations between environmental events. In short, the connectivity of the sensory association cortex is altered to permit the mediation of environment–environment relations.

Thus far, the hippocampal-derived diffuse projection system selects connections in the sensory association cortex to mediate environment–environment relations, and the VTA-derived diffuse projection system selects connections in the motor association cortex to mediate environment–behavior relations. The selection of these two types of relations must be coordinated so that the correlation between environmental events is most appreciated at those times when responses are followed by reinforcers. How is this accomplished? At the neural level, coordination is implemented by axons from VTA neurons that project to synapses of CA1 hippocampal neurons (Swanson, 1982). Dopamine from these axons is known to increase the ability of CA1 cells to be driven by their inputs (Stein & Belluzzi, 1988, 1989). Thus when responses are followed by reinforcers, the diffuse hippocampus-derived neuromodulatory signal is strongest and environment–environment relations preceding the responses are most rapidly selected. For example, if the co-occurrence of auditory and visual stimuli reliably precede a reinforced response, the motor association cortex would be provided with inputs whose activity signaled the co-occurrence of auditory and visual stimuli (AV) as well as the separate occurrences of auditory (A) and visual (V) stimuli (Figure 3). The implications of this account for the interpretation of phenomena such as "perceptual learning," "latent learning," the formation of equivalence classes, and phoneme-grapheme correspondences in verbal behavior are described elsewhere (Donahoe & Palmer, in press).

In summary, experimental analyses of behavior and the neurosciences are converging upon a powerful conception of a principle of reinforcement. By means of widely broadcast neural systems for implementing reinforcement, the eligibility for selection is maintained for the full range of stimuli that can be sensed and responses that can be emitted. Organisms equipped with diffuse neural systems for implementing selection by reinforcement appear to be well equipped to acquire complex environment–behavior relations and the environment–environment relations upon which such relations sometimes depend. Tracing the implications of a principle of reinforcement is not a simple task, however, and requires equally powerful techniques of interpretation. We now turn to one such technique, adaptive neural networks (Donahoe & Palmer, 1989).

INTERPRETATION OF REINFORCEMENT VIA SELECTION NETWORKS

Complex behavior is the product of such a prolonged history of selection that experimental analysis is often precluded. Faced with impediments to experimental analysis, other historical sciences have supplemented experimental analysis with interpretation. Scientific interpretation differs from mere speculation in that interpretation makes use only of principles that are derived from independent experimental analyses. New principles are never uncovered through interpretation, although interpretation may reveal new implications of existing principles (cf. Donahoe & Palmer, 1989, in press).

The interpretive technique used here to explore the implications of a reinforcement principle is computer simulation via adaptive neural networks (cf. Donahoe & Palmer, 1989; Kehoe, 1989). An adaptive neural network is an interconnected set of units whose characteristics are constrained by findings from experimental analyses of the neurosciences. If biobehaviorally constrained computer simulations yield results that are consistent with empirical observations and do not yield inconsistent results, then the principles that inform the simulations are accepted (with the tentativeness accompanying all conclusions in science) as explanations of the phenomena. Adaptive neural networks are not "conceptual nervous systems" in Skinner's sense because their characteristics are constrained by exper-

imental analyses of directly observed structures and processes. Their characteristics are not the result of inferences from observations at one level (the behavioral level) to processes at another level (the neural level).

The adaptive neural networks used in the following simulations contain input units that simulate the state of the environment, interior (often called "hidden") units, and output units that simulate the behavior of the network. When environmental events occur, they activate input units; this activation probabilistically activates other interior or output units to which they are connected. The units and their connections define the architecture of the network. In the simulations described here, the architecture and functioning of the reinforcement system reflect what is known about diffuse neural systems in modifying synaptic efficacies, or connection weights, between units.[2] Because the cumulative action of the diffuse reinforcement systems selects those connections that most reliably mediate reinforced input–output relations, we call such networks *selection networks* (Donahoe & Palmer, in press). A selection network is composed of two subnetworks: (a) a response-selection component that mediates environment–behavior relations and simulates the VTA-motor cortex system and (b) a stimulus-selection component that mediates environment–environment relations and simulates the hippocampal-sensory cortex system. The output of the stimulus-selection component serves as input to the response-selection component and thereby enriches the stimulus patterns guiding behavior. The main architectural features of a selection network are shown in Figure 4.

The functioning of a selection network may be summarized as follows. Simulated environmental events activate the input units of the stimulus-selection component. Within the stimulus-selection component, co-occurrences of environmental events selectively strengthen connections between simultaneously activated input and polysensory units through simulation of the effect of the diffusely projecting hippocampal system. The selection of pathways that mediate environment–environment relations proceeds more rapidly when co-occurrences are followed by reinforcers that activate the simulated VTA system and amplify the diffuse hippocampal signal. Within the response-selection component, reinforcers selectively strengthen connections between inputs coming from the stimulus-selection component and simultaneously active units in the response-selection component through simulation of the diffusely projecting VTA system. The cumulative effect of these coordinated selection processes forges functional pathways through the selection network that mediate relations between events that are reliably present prior to reinforcers. (Equations defining the activation rule, whereby one unit activates another, and the learning rules, whereby the connection weights between units are modified, are given in the Appendix.)

Simulation of Classical and Operant Contingencies

The unified reinforcement principle, when simulated in a selection network, should enable a network to mediate the environment–behavior relations characteristically produced by respondent and operant contingencies. Figure 5 depicts two response-selection components of identical architecture but with different contingencies for the presentation of the reinforcer. (The simulated procedures do not require the selection of environment–environment relations because the inputs to the stimulus-selection component are sufficient to distinguish environments in which reinforcers occur from those in which they do not. Accordingly, except where indicated, the simulations that follow use the response-selection component only.) In the upper network, a respondent (or classical) procedure is illustrated: The reinforcing stimulus (the unconditioned stimulus, or US) is contingent upon an environmental event (the conditioned stimulus, or S_1). The respondent contingency is indicated by the broken line. In the lower network, an operant procedure is illustrated: In the presence of a given environment, S_1, the reinforcing

---

[2] The remaining architectural features are simplified in these simulations because the focus is upon the reinforcement system. In more complete simulations, the architecture is constrained by other neuroanatomical research or is the product of the simulation of neural development by means of what are called *genetic algorithms* (e.g., Davis, 1991). A genetic algorithm simulates the effect of natural selection on network architecture, whereas a learning algorithm, which is the focus here, simulates the effect of selection by reinforcement on synaptic efficacies.
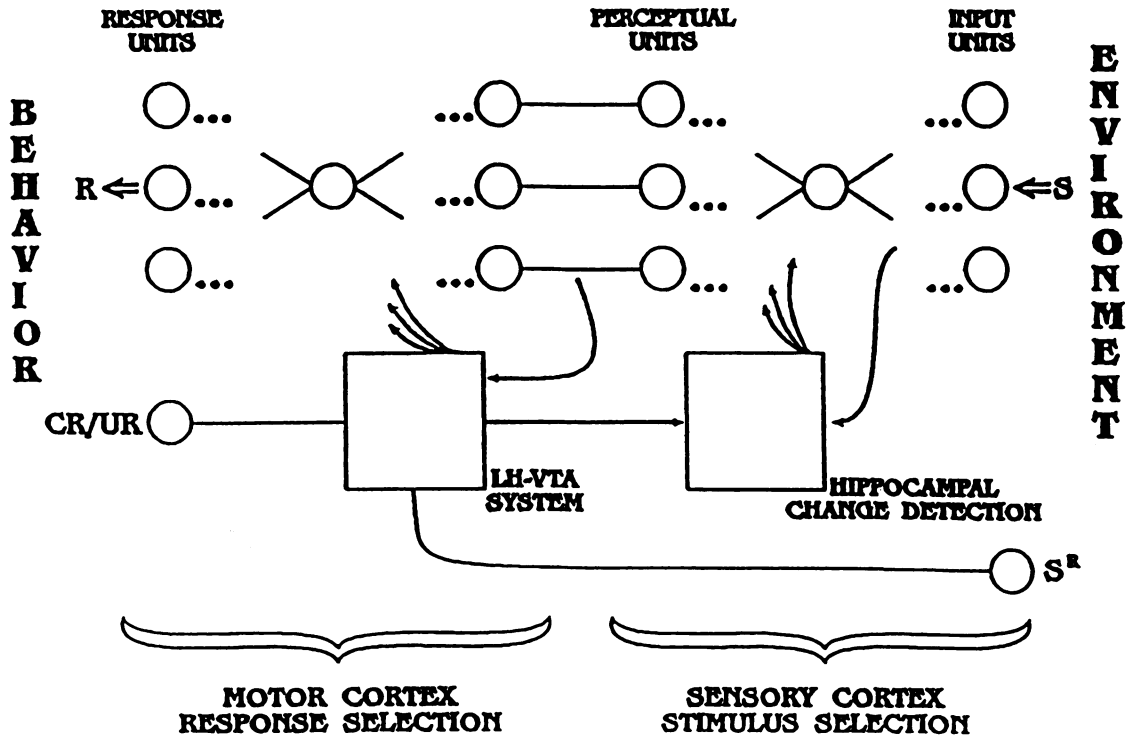
SELECTION NETWORK

Fig. 4.   The architecture of a selection network, an adaptive neural network whose structure simulates the neural systems mediating reinforcement in the brain. Shown are the diffusely projecting lateral hypothalamic (LH) and ventral tegmental area (VTA) systems that select pathways mediating environment–behavior relations and the diffusely projecting hippocampal system that selects pathways mediating environment–environment relations.

stimulus (here also symbolized by US) is contingent on a behavioral event (activation of the operant unit, or R). The broken line now indicates an operant contingency.

In both simulations, whenever a reinforcing stimulus evoked an increase in the elicited response (an increase in the activation of the UR/CR unit), a reinforcing signal that was proportional to the increase was broadcast throughout the network. Momentary increases in the activation of the UR/CR unit strengthened connection weights between *all* active units, whether in the respondent or operant procedure. (Simulations in which connection weights are modified, or "updated," several times within a "trial" are known as "real-time" simulations. All simulations reported here are real-time simulations in which the connection weights were updated 10 times per

trial and any reinforcer occurred on the 10th time step. Multiple updates within a trial simulate the continuously changing activity that ensues upon the presentation of a stimulus and the propagation of that activity through an interconnected set of neurons.)

In the classical procedure, the coactivated units, whose connections are the only ones eligible for strengthening at a given moment, necessarily include those interior units activated by the conditioned stimulus (S_1) input unit and the output unit simulating the unconditioned (and conditioned) response (UR/CR unit). In the operant procedure, in addition to connections from S_1 to interior units, some of the strengthened connections are necessarily from interior units to the output unit simulating the operant, R. The R unit must be activated on any trail in which a reinforcer
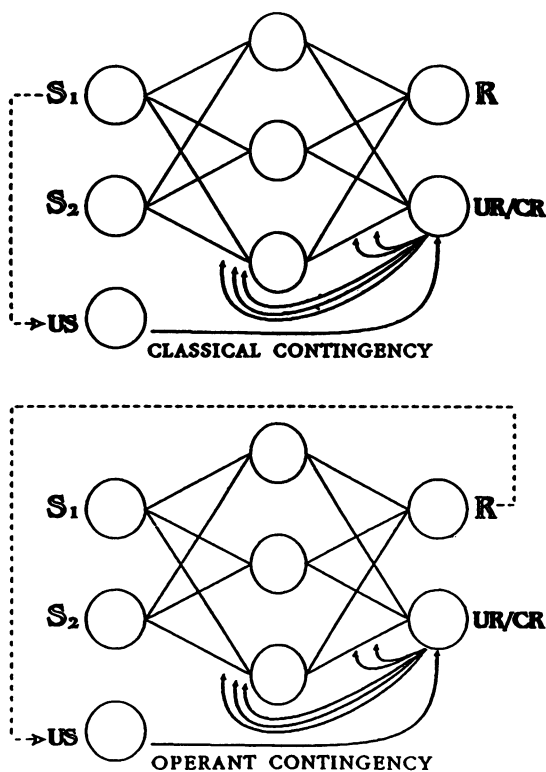
Fig. 5. Response-selection component of a selection network showing networks having three input units, three interior units, and two output units. The input units can be activated by either of two environmental stimuli (S_1 and S_2) and the putative reinforcing stimulus (US). Activation of the output units simulates the occurrence of the operant response (R) and the unconditioned response (UR) or, when the UR unit is activated by a stimulus other than the US, the conditioned response (CR). Activation of the UR/CR unit also engages the diffuse reinforcing system that modifies connection weights between all coactive units. The upper panel illustrates a classical, or respondent, contingency in that the occurrence of the US is dependent on the prior activation of an input unit, S_1. The lower panel depicts an operant, or instrumental, contingency in that activation of the US unit is dependent on the prior activation of the output unit, R, by the presentation of S_1.

occurs, because activation of the operant unit is required by the operant contingency.

Acquisition. Computer simulations of classical and operant procedures were conducted

in which selection occurred as described by the unified reinforcement principle. The same parameter values (see Appendix) were used for all simulations. (Qualitative aspects of the findings reported here are not restricted to either the specific network architectures or parameter values used in the simulations, although both affect quantitative aspects of the findings and are critical in the simulation of some phenomena.) Examples of the outcomes of these simulations are shown in Figure 6. The leftmost upper panel depicts acquisition of the CR in the classical procedure, where the measure of conditioning is the level of activation of the UR/CR, or respondent, unit by the S_1. As can be seen, the respondent unit was initially unreliably activated at very low levels by the S_1 but, after some 30 US presentations, the broadcast reinforcement signal had sufficiently strengthened connections along pathways between the S_1 and US/CR units for the network to mediate the S_1–CR relation strongly and reliably.

Acquisition with an operant contingency is shown in the leftmost lower panel of Figure 6. Here, the occurrence of the reinforcer (i.e., activation of the US unit) was contingent on the prior activation of the operant (R) unit. At first, the R unit was rarely and weakly activated by the environmental stimulus (S_1). (Concomitant changes in activation of the CR unit are discussed shortly.) This initial level of activation corresponds to the baseline or operant level of the R unit. Whenever the R unit was activated, the reinforcer was presented and the diffuse reinforcement system was activated. The connection weights between all coactive units were then strengthened. As shown in Figure 6, the cumulative effect of this process was that connection weights were strengthened along pathways that permitted the network to mediate the S_1–R relation after some 50 presentations of the reinforcer. Thus the unified reinforcement principle implemented in a selection network is competent to simulate acquisition with an op-

→

Fig. 6. Simulation of acquisition, extinction, and reacquisition with a respondent contingency (upper panel) and an operant contingency (lower panel). The activation levels of the CR output unit are shown for the classical contingency, and the activation levels of the R and CR output units are shown for the operant contingency over a number of simulated conditioning trials. The simulations were real-time simulations (see text) in which the connection weights were adjusted at 10 time steps (t) within each trial. The activation levels at Time-Step 9, which is the time step before the occurrence of any reinforcer, are shown.
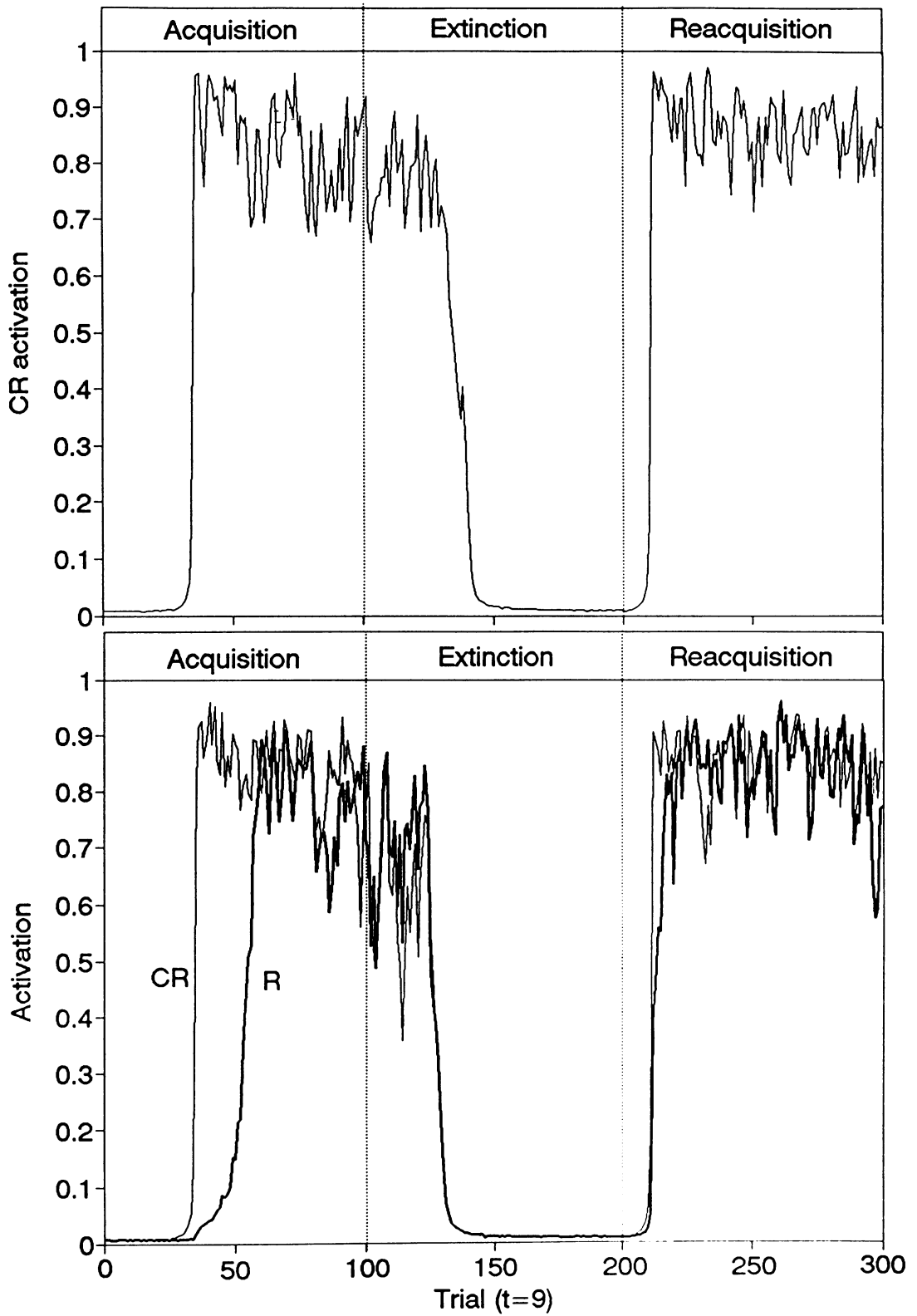
erant as well as with a respondent contingency. Respondent and operant procedures do not necessarily require different principles for their interpretations. Instead, the different outcomes of the two procedures may be viewed as the emergent and cumulative products of the same principle.

Extinction and reacquisition. The middle sections of Figure 6 show the simulated effects of extinction in the respondent (upper panel) and operant (lower panel) procedures. Activation of a unit without the concurrent activation of the diffuse reinforcement system decreased the connection weights between all coactive units. As the cumulative effect of this process, stimulation of the S_1 unit gradually lost its ability to activate the CR unit in the respondent procedure and the R unit in the operant procedure. An emergent effect of the simulation of extinction is illustrated in the rightmost sections of both panels: When the US (reinforcer) unit was again activated in accordance with respondent (upper panel) or operant (lower panel) contingencies, S_1 *more rapidly* reacquired its ability to activate the CR and R units, respectively, than in original acquisition. Reacquisition was facilitated by the following process: During extinction, the repeated activation of S_1 without reinforcement weakened connection weights from the S_1 unit to interior units most rapidly, because these units were most frequently activated. Once connections from S_1 to interior units had weakened sufficiently for the interior units to be no longer activated, connection weights from interior units to the output units were "protected" from further weakening. (Only connection weights of activated units may change.) Thus, during the simulation of extinction, connection weights to units "deep" within the network remained relatively unchanged. Then, when reacquisition began, these intact connections weights were available to facilitate reconditioning (cf. Kehoe, 1988).

Simulation of Acquired Reinforcement

Some of the potential contributions of acquired reinforcement to complex behavioral chains were noted earlier. Computer simulation also reveals important emergent contributions of the neural mechanisms of acquired reinforcement even during simple conditioning. The lower panels of Figure 6 depict the activation levels of the respondent (CR) unit as well as the operant (R) unit during operant conditioning, extinction, and reconditioning. According to the unified reinforcement principle, respondents and operants are concurrently acquired when an operant contingency is implemented.

First, note that activation of the CR unit by S_1 was acquired *before* activation of the R unit. This is a general result and occurs for two primary reasons: (a) The CR unit is more strongly activated by the US from the outset of conditioning than is the R unit by S_1 (i.e., the respondent is elicited by the US, whereas the operant is emitted in the presence of S_1). (b) The delay in reinforcement after activating the R unit is necessarily greater than after activating the CR/UR unit. (The reinforcer— activation of the US unit—necessarily occurs *after* activation of the R unit in an operant procedure, whereas the diffuse reinforcing signal occurs immediately upon activation of the CR/UR unit.) Because changes in connection weights are directly related to the activation levels of the coactive units and because activation levels decay over time, changes in connection weights to the CR/UR unit occur more rapidly than to the R unit.

The more rapid acquisition of the CR than R in operant procedures has implications for the interpretation of a number of behavioral phenomena. Only two are considered here. First, when the conditioned response is incompatible with the operant, the putative reinforcer will be relatively ineffective for that operant, because the CR gains strength before the R. In this way, many so-called biological constraints on learning may be seen as emergent outcomes of the reinforcement principle itself (Breland & Breland, 1961; cf. Donahoe et al., 1982). Second, the same CR–R interactions provide insight into the punishment procedure. If the putative punisher elicits responses (e.g., withdrawal) that are incompatible with the operant, then the more rapid acquisition of the CR prevents R from gaining strength. If, however, an aversive stimulus elicits responses that are compatible with the operant, then the aversive stimulus will function as a *reinforcer* for that operant (cf. Kelleher & Morse, 1968). Other phenomena, such as devaluation (Rescorla, 1991) and autoshaping (e.g., D. Williams & Williams, 1969), may also be interpreted as dependent upon the more rapid acquisition of the CR than R and are

discussed elsewhere (Donahoe & Palmer, in press).[3]

A second important consequence of the conditioning of CR activity in the operant procedure is the effect on acquired reinforcement: Because activation of the CR/UR unit engages the diffuse reinforcement system, activation of the CR/UR unit by S_1-initiated activity implements acquired reinforcement *within a trial during simple conditioning*. Some sense of the effect of acquired reinforcement on conditioning is shown during acquisition in the lower left panel of Figure 6. Trials during which the CR unit was strongly activated by S_1 were generally followed by increases in the ability of S_1 to initiate activity in the R unit. The operation of acquired reinforcement within the network strengthened the S_1–R relation. Thus the acquired reinforcement system mediates what might loosely be called "self-reinforcement" (cf. Catania, 1975). During extinction (see lower middle panel of Figure 6), periods of transiently increased activation of the R unit following increases in activation of the CR unit may reveal the mechanism responsible for changes in response strength that are labeled as "emotional" (Skinner, 1938). The effects of acquired reinforcement facilitate acquisition, particularly in deeply layered networks (cf. Donahoe & Palmer, in press), and retard extinction in operant conditioning. In short, acquired reinforcement operating within a selection network has emergent effects on the behavioral changes produced by an operant contingency.

Simulation of Stimulus-Control Phenomena

The final set of simulations uses selection networks to demonstrate three phenomena of stimulus control—blocking, stimulus discrimination, and stimulus configuring.

Blocking. Blocking was simulated using a respondent contingency in which activation of the CS_1 unit was first paired with activation of the US unit until CS_1 controlled the CR/UR unit (see the left panel of Figure 7). Then simulation of CS_1-US pairing continued, but with the CS_2 unit now activated concurrently with the CS_1 unit. As shown in the middle panel of Figure 7, the CR unit was reliably activated by the co-occurrence of the compound CS of CS_1 and CS_2. To assess the ability of CS_2 alone to control the CR unit after 100 pairings with the compound CS, the CS_2 unit was activated by itself to simulate probe tests. As shown in the right panel of Figure 7, the ability of CS_2 to activate the CR unit was blocked. Blocking occurred because (a) the reinforcer-induced behavioral discrepancy was reduced at the outset of the second phase of conditioning (CS_1 already activated the CR/UR unit prior to presentation of the US) and (b) the pathways activated by CS_1 and CS_2 were in competition with one another for control of the CR/UR unit. (The relative contribution of these two processes to blocking depends upon the architecture of the network.) Thus blocking of the acquisition of stimulus control may be simulated by a selection network.

Stimulus discrimination. The acquisition of an operant intradimensional discrimination (Honig, 1970) was simulated using a selection network with three input units—S_1, S_2, and S_3. (An intradimensional procedure is one in which the stimuli to be discriminated differ within the same sensory dimension, e.g., different wavelengths of light or frequencies of tone.) The stimulus during which activation of the R unit was reinforced (i.e., S+) consisted of the simultaneous activation of the S_1 and S_2 input units. The stimulus during which activation of the R unit was nonreinforced (i.e., S−) consisted of the simultaneous activation of the S_2 and S_3 input units. Thus the S_2 input unit was activated during both S+ and S− to simulate the common receptors activated by

---

[3] It would be a mistake to view the foregoing discussion of the acquisition of CR and R with an operant contingency as indicating that the unified reinforcement principle is some variant of two-factor theory (Mowrer, 1947; Rescorla & Solomon, 1967). To the contrary, the present view argues against multifactor accounts of conditioning and for a single reinforcement principle that has different cumulative effects in respondent and operant procedures. That is, there is only one factor—conditioning. Similarly, it would be a mistake to view the present formulation as an attempt to "reduce" operant conditioning to respondent conditioning (e.g., Commons, Bing, Griffy, & Trudeau, 1991; Trapold & Overmier, 1972; cf. Pear & Eldridge, 1984). Again, there is only one fundamental process—conditioning. Further, conditioning in whatever procedure it occurs can be accommodated by a single principle that produces quite different behavioral outcomes depending upon which stimuli and responses are reliably contiguous with a behavioral discrepancy and the compatibility of those responses. Historically, the view that is closest to the present conception is that of Hilgard and Marquis (1940), who tentatively regarded the operant-respondent distinction as procedural and capable, perhaps, of being interpreted by a common principle of conditioning.
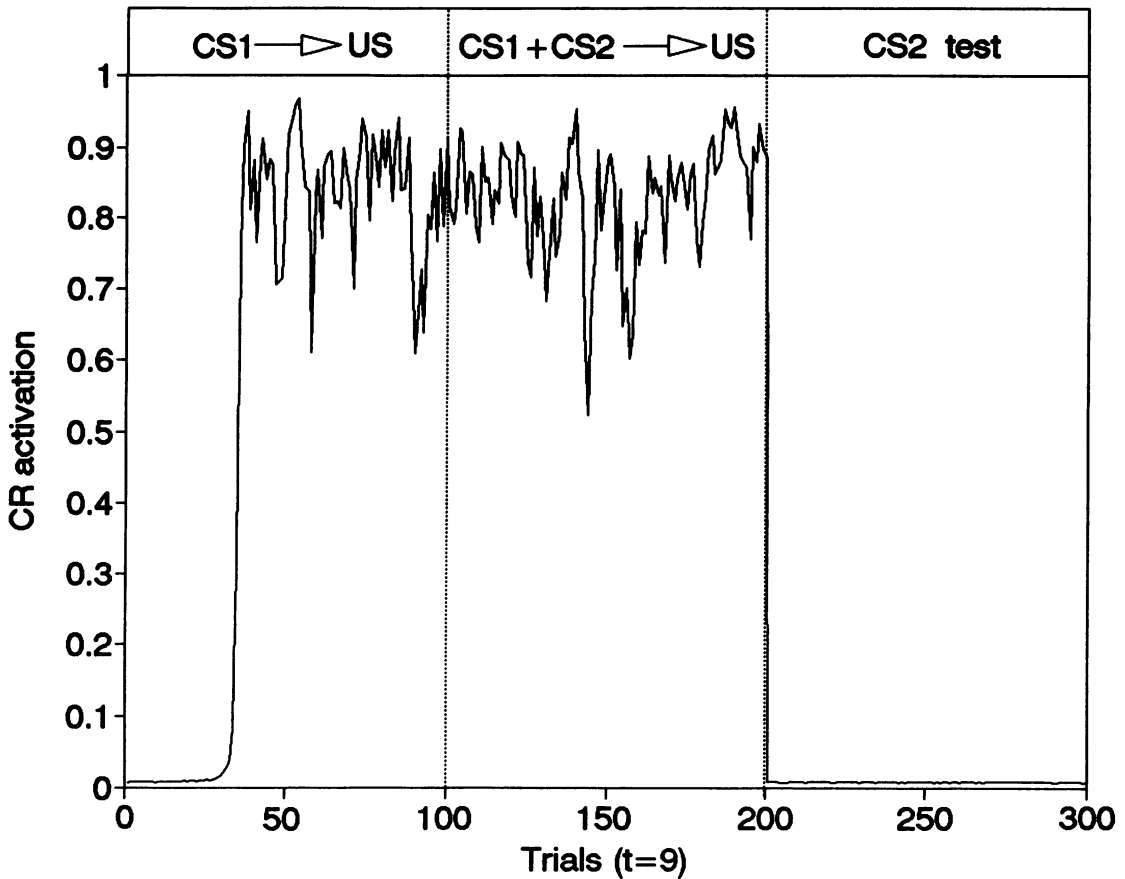
Fig. 7.   Simulation of blocking in a three-phase classical conditioning experiment. During the first phase, activation of the US input unit was contingent on the prior activation of the CS1 input unit. During Phase 2, activation of the US unit was contingent on the activation of both the CS1 and CS2 units. During the test phase, only the CS2 unit was activated, with the learning algorithm disengaged for "probe" tests to determine whether CS2 had acquired control over the CR unit.

similar stimuli within the same sensory dimension. (In these simulations, S_2 was more strongly activated than either S_1 or S_3 to increase the similarity between S+ and S−.) As shown in Figure 8, S+ acquired strong control over the R unit, whereas S− only weakly controlled the R unit after a period during which the activation level increased (cf. Hanson, 1959).

Stimulus configuring. The last simulation illustrates the functioning of the stimulus-selection component of a selection network. Suppose that a response is reinforced following the coactivation of input units S_1 and S_2 but is nonreinforced when either S_1 or S_2 is activated alone. Under these conditions, the stimulus-

selection component is capable of differentially strengthening the connection weights from the S_1 and S_2 units to an interior polysensory unit with the result that the co-occurrence of S_1 and S_2 reliably activates the polysensory unit. That is, the stimulus-selection component "constructs" a polysensory S_{12} unit that can then control activation of the R unit.

Figure 9 depicts the activation levels of a polysensory unit within the stimulus-selection component when reinforcers occurred only after the S_1 and S_2 units were simultaneously activated. One simulation shows the activation levels of the polysensory unit when the simulated reinforcement signal from the response-selection component was large (US = 0.9); the
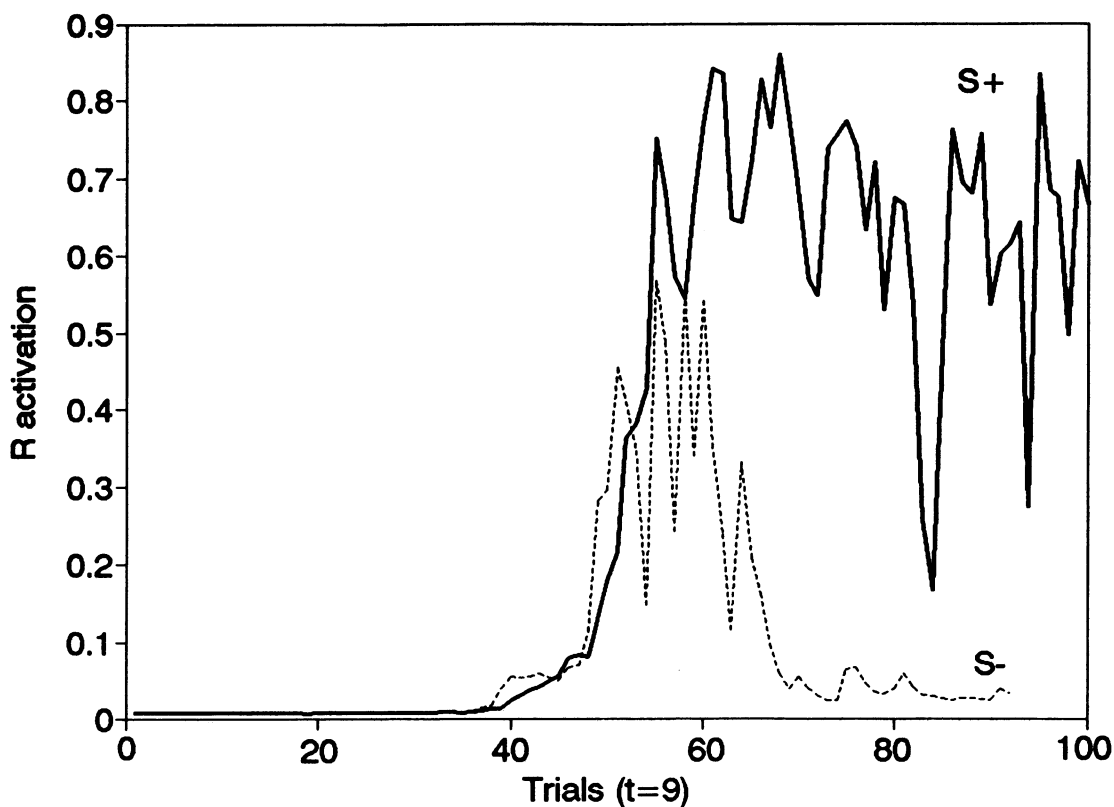
Fig. 8. Simulation of the acquisition of an intradimensional operant stimulus discrimination by a selection network. Shown are the activation levels of the operant (R) unit initiated by the positive stimulus (S+) and the negative stimulus (S−). S+ activated two input units (S₁ and S₂); if the R output unit became active, then the US input unit, which functioned as the reinforcer, was activated. S− activated two input units (S₂ and S₃) but, whether or not the R unit became active, the reinforcer was never presented.

other shows the activation levels when the reinforcement signal was small (US = 0.1). Initially, the polysensory unit was only weakly activated by the co-occurrence of S_1 and S_2 but ultimately became strongly activated, with the increase occurring more rapidly for the larger reinforcement signal. Thus the stimulus-selection component has the ability to strengthen connections, on-line and "as needed," to polysensory units whose activity reflects the environment–environment relations detected by the input units, particularly when those relations occur prior to reinforcers. The implications of the stimulus-selection component for the interpretation of such complex phenomena as place learning (O'Keefe & Nadel, 1978), latent learning (Tolman, 1932), declarative versus procedural "memory" (Squire, 1992), and the formation of equivalence classes (Sidman

& Tailby, 1982) are discussed elsewhere (Donahoe & Palmer, in press).

CONCLUDING COMMENTS

A principle of reinforcement has been described whose formulation is constrained by experimental analyses of both behavior and the neurosciences. This principle—the unified reinforcement principle—is competent to yield many of the basic phenomena produced by operant and respondent contingencies, including acquired reinforcement and stimulus control, when implemented in a class of adaptive neural networks known as selection networks.

From a selectionist perspective, complex phenomena are emergent outcomes of the cumulative action of relatively simple processes. Whether the unified reinforcement principle
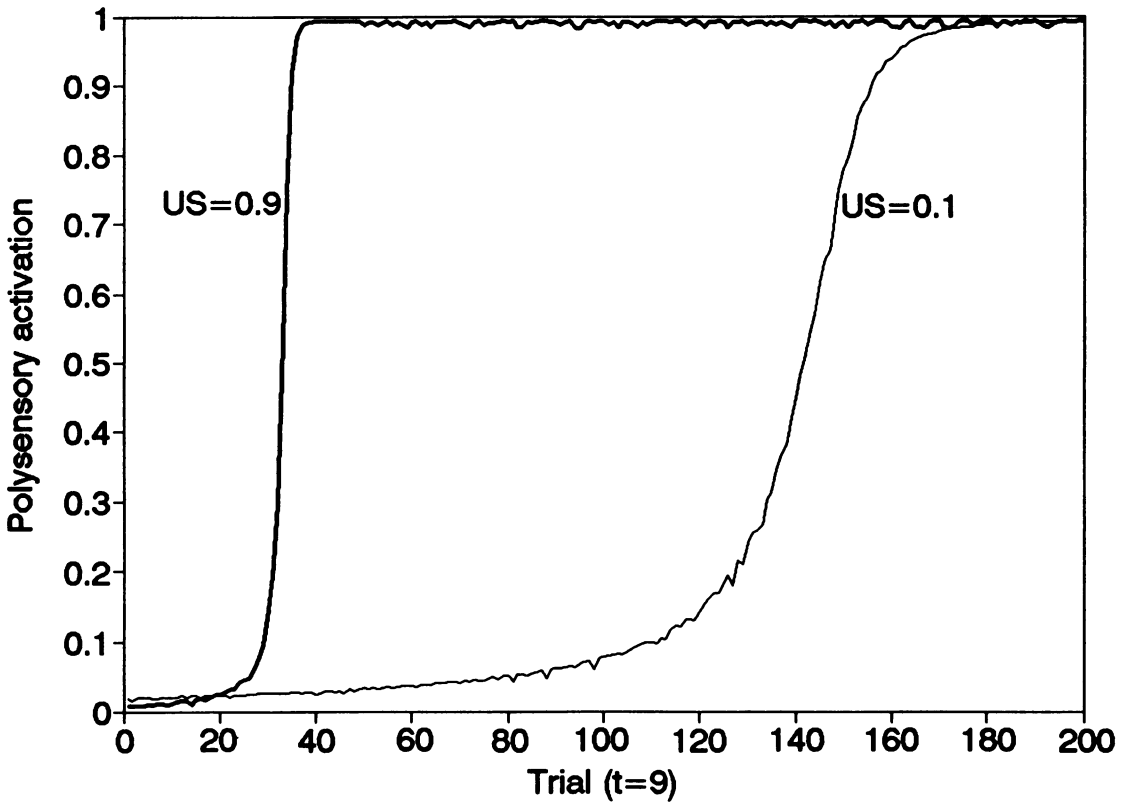
Fig. 9. Simulations of the strengthening of connections to a polysensory unit within the stimulus-selection component of a selection network. Shown are the activation levels of a polysensory unit as a function of the number of concurrent activations of two input units, S_1 and S_2. In one simulation, the magnitude of the simulated VAT reinforcing signal was low (US = 0.1). In the other simulation, the magnitude was high (US = 0.9).

will fulfill its promise as a general formulation for describing the selecting effect of the individual environment on behavior awaits further experimental analyses at the behavioral and neural levels. Whether it will prove sufficiently powerful to yield truly complex behavior when implemented in neural networks of more complex architectures awaits further simulation research. No principled impediments appear to exist to either enterprise.

REFERENCES

Amaral, D. G. (1987). Memory: Anatomical organization of candidate brain regions. In F. Plum (Ed.) *Handbook of physiology: Sec. 1. Neurophysiology: Vol. 5. Higher functions of the brain* (pp. 211–294). Bethesda, MD: American Physiological Society.

Beninger, R. J. (1983). The role of dopamine activity in locomotor activity and learning. *Brian Research Reviews,* **6,** 173–196.

Bliss, T. V. P., & Lomo, T. (1973). Long-lasting po-tentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. *Journal of Physiology (London),* **232,** 331–356.

Bowler, P. J. (1983). *The eclipse of Darwinism: Anti-Darwinian evolution theories in the decades around 1900.* Baltimore, MD: Johns Hopkins University Press.

Breland, K., & Breland, M. (1961). The misbehavior of organisms. *American Psychologist,* **16,** 681–684.

Campbell, D. T. (1974). Evolutionary epistemology. In P. A. Schlipp (Ed.), *The library of living philosophers: Vol. 14-1. The philosophy of Karl Dopper* (pp. 413–463). LaSalle, IL: Open Court Publishing Co.

Catania, A. C. (1975). The myth of self-reinforcement. *Behaviorism,* **3,** 192–199.

Catania, A. C. (1987). Some Darwinian lessons for behavior analysis: A review of Bowler's *The Eclipse of Darwinism. Journal of the Experimental Analysis of Behavior,* **47,** 249–257.

Coleman, S. R. (1981). Historical context and systematic functions of the concept of the operant. *Behaviorism,* **9,** 207–226.

Coleman, S. R. (1984). Background and change in B. F. Skinner's metatheory from 1930 to 1938. *Journal of Mind and Behavior,* **5,** 471–500.

Commons, M. L., Bing, E. W., Griffy, C. C., & Trudeau, E. J. (1991). Models of acquisition and preference. In M. L. Commons, S. Grossberg, & J. E. R. Staddon (Eds.), *Neural network models of conditioning and action* (pp. 201–223). Hillsdale, NJ: Erlbaum.

Davis, L. (Ed.). (1991). *Handbook of genetic algorithms*. New York: Van Nostrand Reinhold.

Dinsmoor, J. A. (1950). A quantitative comparison of the discriminative and reinforcing functions of a stimulus. *Journal of Experimental Psychology,* **40,** 458–472.

Dobzhansky, T. G. (1937). *Genetics and the origin of species.* New York: Columbia University Press.

Donahoe, J. W., Crowley, M. A., Millard, W. J., & Stickney, K. A. (1982). A unified principle of reinforcement: Some implications for matching. In M. L. Commons, R. J. Herrnstein, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 2. Matching and maximizing accounts* (pp. 493–521). Cambridge, MA: Ballinger.

Donahoe, J. W., & Palmer, D. C. (1989). The interpretation of complex human behavior: Some reactions to *Parallel Distributed Processing,* edited by J. L. McClelland, D. E. Rumelhart, and the PDP Research Group. *Journal of the Experimental Analysis of Behavior,* **51,** 399–416.

Donahoe, J. W., & Palmer, D. C. (in press). *Learning and complex behavior.* Boston: Allyn & Bacon.

Fallon, J. H., & Loughlin, S. E. (1987). Monoamine innervation of cerebral cortex and a theory of the role of monoamines in cerebral cortex and basal ganglia. In E. G. Jones & A. Peters (Eds.), *Cerebral cortex: Vol. 6. Further aspects of cortical function, including hippocampus* (pp. 41–127). New York: Plenum.

Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement.* New York: Appleton-Century-Crofts.

Gehrz, R. D., Black, D. C., & Solomon, P. M. (1984). The formation of stellar systems from interstellar molecular clouds. *Science,* **224,** 823–830.

Gimpl, M. P., Gormezano, I., & Harvey, J. A. (1979). Effect of haloperidol and pimozide (PIM) on Pavlovian conditioning of the rabbit nictitating membrane response. In E. Usdin, I. J. Kopin, & J. Barchas (Eds.), *Catecholamines: Basic and clinical frontiers* (Vol. 2, pp. 1711–1713). New York: Pergamon Press.

Hanson, H. M. (1959). Effects of discrimination training on stimulus generalization. *Journal of Experimental Psychology,* **58,** 321–334.

Herrnstein, R. J. (1970). On the law of effect. *Journal of the Experimental Analysis of Behavior,* **13,** 243–266.

Herrnstein, R. J. (1982). Melioration as behavioral dynamism. In M. L. Commons, R. J. Herrnstein, & H. Rachlin (Eds.), *Quantitative analyses of behavior: Vol. 2. Matching and maximizing accounts* (pp. 433–458). Cambridge, MA: Ballinger.

Heth, C. D. (1992). Levels of aggregation and the generalized matching law. *Psychological Review,* **99,** 306–321.

Hilgard, E. R., & Marquis, D. G. (1940). *Conditioning and learning.* New York: Appleton-Century.

Hinson, J. M., & Staddon, J. E. R. (1983). Hill-climbing by pigeons. *Journal of the Experimental Analysis of Behavior,* **39,** 25–47.

Hoebel, B. G. (1988). Neuroscience and motivation: Pathways and peptides that define motivational systems. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology: Vol. 1. Perception and motivation* (pp. 547–625). New York: Wiley.

Honig, W. K. (1970). Attention and the modulation of stimulus control. In D. I. Mostofsky (Ed.), *Attention: Contemporary theory and analysis* (pp. 193–238). New York: Appleton-Century-Crofts.

Hull, D. L. (1973). *Darwin and his critics; the reception of Darwin's theory of evolution by the scientific community.* Cambridge, MA: Harvard University Press.

Irikl, A., Pavlides, C., Keller, A., & Asanuma, H. (1989). Long-term potentiation in the motor cortex. *Science,* **245,** 1385–1387.

Kamin, L. J. (1968). "Attention-like" processes in classical conditioning. In M. R. Jones (Ed.), *Miami symposium on the prediction of behavior* (pp. 9–31). Coral Gables, FL: University of Miami Press.

Kamin, L. J. (1969). Predictability, surprise, attention and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296). New York: Appleton-Century-Crofts.

Kehoe, E. J. (1988). A layered network model of associative learning: Learning to learn and configuration. *Psychological Review,* **95,** 411–433.

Kehoe, E. J. (1989). Connectionist models of conditioning: A tutorial. *Journal of the Experimental Analysis of Behavior,* **52,** 427–440.

Kelleher, R. T., & Morse, W. H. (1968). Schedules using noxious stimuli: III. Responding maintained by response-produced electric shocks. *Journal of the Experimental Analysis of Behavior,* **11,** 819–838.

Keller, F. S., & Schoenfeld, W. N. (1950). *Principles of psychology: A systematic text in the science of behavior.* New York: Appleton-Century-Crofts.

Kety, S. S. (1970). The biogenic amines in the central nervous system: Their possible role in arousal, emotion, and learning. In F. O. Schmidtt (Ed.), *Neuroscience second study program* (pp. 324–336). New York: Rockefeller University Press.

Krieckhaus, E. E., Donahoe, J. W., & Morgan, M. A. (1992). Paranoid schizophrenia may be caused by dopamine hyperactivity of CA1 hippocampus. *Biological Psychiatry,* **31,** 560–570.

Levy, W. B., & Desmond, N. L. (1985). The rules of elemental synaptic plasticity. In W. B. Levy, J. A. Anderson, & S. Lehmkuhle (Eds.), *Synaptic modification, neuron selectivity, and nervous system organization* (pp. 105–121). Hillsdale, NJ: Erlbaum.

Long, J. B. (1966). Elicitation and reinforcement as separate stimulus functions. *Psychological Reports,* **19,** 759–764.

Lowel, S., & Singer, W. (1992). Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity. *Science,* **255,** 209–212.

Mackintosh, N. J. (1983). *Conditioning and associative learning.* New York: Oxford University Press.

Marshall-Goodell, B., & Gormezano, I. (1991). Effects of cocaine on conditioning of the rabbit nictitating membrane response. *Pharmacology Biochemistry and Behavior,* **39,** 503–507.

Maynard Smith, J. (1982). *Evolution and the theory of games.* Cambridge: Cambridge University Press.

Mayr, E. (1982). *The growth of biological thought: Diversity, evolution, and inheritance.* Cambridge, MA: Belknap Press.

Morse, W. H. (1966). Intermittent reinforcement. In W. K. Honig (Ed.), *Operant behavior: Areas of research*

and application (pp. 52–108). New York: Appleton-Century-Crofts.

Morse, W. H., & Skinner, B. F. (1957). A second type of superstition in the pigeon. *American Journal of Psychology, 70,* 308–311.

Mowrer, O. H. (1947). On the dual nature of learning: A reinterpretation of "conditioning" and "problem-solving." *Harvard Educational Review, 17,* 102–150.

Nevin, J. A. (1979). Overall matching versus momentary maximizing: Nevin (1969) revisited. *Journal of Experimental Psychology: Animal Behavior Processes, 5,* 300–306.

O'Keefe, J., & Nadel, L. (1978). *The hippocampus as a cognitive map.* Oxford: Clarendon Press.

Palmer, D. C. (1987). *The blocking of conditioned reinforcement.* Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Palmer, D. C., & Donahoe, J. W. (1992). Essentialism and selection in cognitive science and behavior analysis. *American Psychologist, 47,* 1344–1358.

Pear, J. J., & Eldridge, G. D. (1984). The operant-respondent distinction: Future directions. *Journal of the Experimental Analysis of Behavior, 42,* 453–467.

Rescorla, R. A. (1969). Conditioned inhibition of fear. In N. J. Mackintosh & W. K. Honig (Eds.), *Fundamental issues in associative learning* (pp. 65–89). Halifax, Nova Scotia: Dalhousie University Press.

Rescorla, R. A. (1991). Associative relations in instrumental learning: The eighteenth Bartlett memorial lecture. *Quarterly Journal of Experimental Psychology, 43B,* 1–23.

Rescorla, R. A., & Solomon, R. L. (1967). Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. *Psychological Review, 74,* 151–182.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 318–362). Cambridge, MA: MIT Press.

Schoenfeld, W. N., Cole, B. K., Blaustein, J., Lachter, G. D., Martin, J. M., & Vickery, C. (1972). *Stimulus schedules: The t-τ systems.* New York: Harper & Row.

Shimp, C. P. (1969). Optimal behavior in free-operant experiments. *Psychological Review, 76,* 97–112.

Shizgal, P., Bielajew, C., & Rompre, P-P. (1988). Quantitative characteristics of the directly stimulated neurons subserving self-stimulation of the medial forebrain bundle: Psychophysical inference and electrophysiological measurement. In M. L. Commons, R. M. Church, J. R. Stellar, & A. R. Wagner (Eds.), *Quantitative analyses of behavior: Vol. 7. Biological determinants of reinforcement* (pp. 59–85). Hillsdale, NJ: Erlbaum.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior, 37,* 5–22.

Silberberg, A., Hamilton, B., Ziriax, J. M., & Casey, J. (1978). The structure of choice. *Journal of Experimental Psychology: Animal Behavior Processes, 4,* 368–398.

Singer, W., & Rauschecker, J. P. (1982). Central core control of developmental plasticity in the kitten visual cortex: II. Electrical activation of mesencephalic and diencephalic projections. *Experimental Brain Research, 47,* 223–233.

Skinner, B. F. (1935a). The generic nature of the concepts of stimulus and response. *Journal of General Psychology, 12,* 40–65.

Skinner, B. F. (1935b). Two types of conditioned reflex and a pseudo type. *Journal of General Psychology, 12,* 66–77.

Skinner, B. F. (1937). Two types of conditioned reflex: A reply to Konorski and Miller. *Journal of General Psychology, 16,* 272–279.

Skinner, B. F. (1938). *The behavior of organisms.* New York: Appleton-Century.

Skinner, B. F. (1948). "Superstition" in the pigeon. *Journal of Experimental Psychology, 38,* 168–172.

Skinner, B. F. (1981). Selection by consequences. *Science, 213,* 501–504.

Squire, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review, 99,* 195–231.

Staddon, J. E. R. (1983). *Adaptive behavior and learning.* Cambridge: Cambridge University Press.

Staddon, J. E. R., & Hinson, J. M. (1983). Optimization: A result or a mechanism? *Science, 221,* 976–977.

Stein, L., & Belluzzi, J. D. (1988). Operant conditioning of individual neurons. In M. L. Commons, R. M. Church, J. R. Stellar, & A. R. Wagner (Eds.), *Quantitative analyses of behavior* (Vol. 7, pp. 249–264). Hillsdale, NJ: Erlbaum.

Stein, L., & Belluzzi, J. D. (1989). Cellular investigations of behavioral reinforcement. *Neuroscience and Biobehavior Reviews, 13,* 69–80.

Stickney, K., & Donahoe, J. W. (1983). Attenuation of blocking by a change in US locus. *Animal Learning & Behavior, 11,* 60–66.

Stone, G. O. (1986). An analysis of the delta rule and the learning of statistical associations. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition: Vol. 1. Foundations* (pp. 444–459). Cambridge, MA: MIT Press.

Swanson, L. W. (1982). The projections of the ventral tegmental area and adjacent regions: A combined fluorescent retrograde tracer and immunofluorescence study in the rat. *Brain Research Bulletin, 9,* 321–353.

Thomas, D. R., & Caronite, S. C. (1964). Stimulus generalization of a positive conditioned reinforcer. II. Effects of discrimination training. *Journal of Experimental Psychology, 68,* 402–406.

Thompson, R. F. (1990). Neural mechanisms of classical conditioning in mammals. *Philosophical Transactions of the Royal Society, London, 161*–170.

Tolman, E. C. (1932). *Purposive behavior in animals and men.* New York: Century.

Trapold, M. A., & Overmier, J. B. (1972). The second learning process in instrumental learning. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning*

II: Current research and theory (pp. 427–452). New York: Appleton-Century-Crofts.

Trowill, J. A., Panksepp, J., & Gandelman, R. (1969). An incentive model of rewarding brain stimulation. *Psychological Review, 76,* 264–281.

Vaughan, W., Jr. (1981). Melioration, matching, and maximizing. *Journal of the Experimental Analysis of Behavior, 36,* 141–149.

vom Saal, W., & Jenkins, H. M. (1970). Blocking the development of stimulus control. *Learning and Motivation, 1,* 52–64.

Williams, B. A. (1990). Enduring problems for molecular accounts of operant behavior. *Journal of Experimental Psychology: Animal Behavior Processes, 16,* 213–216.

Williams, D. R., & Williams, H. (1969). Auto-maintenance in the pigeon: Sustained pecking despite contingent nonreinforcement. *Journal of the Experimental Analysis of Behavior, 12,* 511–520.

Wise, R. A. (1989). The brain and reward. In J. M. Liebman & S. J. Cooper (Eds.), *The neuropharmacological basis of reward* (pp. 377–424). New York: Oxford University Press.

Wise, R. A., & Bozarth, M. A. (1987). A psychomotor stimulant theory of addiction. *Psychological Review, 94,* 469–492.

Woodbury, C. B. (1943). The learning of stimulus patterns by dogs. *Journal of Comparative Psychology, 35,* 29–40.

Yeomans, J. (1988). Mechanisms of brain stimulation reward. In A. N. Epstein & A. R. Morrison (Eds.), *Progress in psychobiology and physiological psychology* (Vol. 13, pp. 227–266). San Diego: Academic Press.

Yeomans, J. S. (1989). Two substrates for medial forebrain bundle self-stimulation: Myelinated axons and dopamine neurons. *Neuroscience and Biobehavioral Reviews, 13,* 91–98.

APPENDIX

Activation Function

Let $N = \{x \in \mathbf{N} \mid 1 \leq x \leq n\}$ and $P = \{j \in N \mid m < j \leq n\}$ be the sets of units and neural processing elements (NPEs) in an artificial neural network, respectively, where $\mathbf{N}$ is the set of positive integers, n is the number of units in the network, and m is the number of input units. Let $R = \{x \in \mathbf{R}^+ \mid 0.0 \leq x \leq 1.0\}$ be the set of possible activation and connection-weight values, where $\mathbf{R}^+$ is the set of positive real numbers. $a: P \times T \rightarrow R$ is the activation function, where $T \subset \mathbf{N}$, the elements of T representing time steps.

The rule for implementing function a in the neural-network simulations is defined as follows. Let $\mathbf{e}(j,t)$ be the vector of activations of the excitatory inputs to postsynaptic element j at t, $\mathbf{i}(j,t)$ be the vector of activations of the inhibitory inputs to j at t, $\mathbf{w}(j,t)$ be the vector of excitatory weights associated with j at t, and $\mathbf{w}'(j,t)$ be the vector of inhibitory weights associated with j at t, where $j \in P$ and $t \in T$. Assuming that $\mathbf{e}(j,t)$, $\mathbf{i}(j,t)$, $\mathbf{w}(j,t)$, $\mathbf{w}'(j,t) \in R^n$, the amounts of excitation (*exc*) and inhibition (*inh*) produced at j during t are given, respectively, by

$$exc(j,t) = \mathbf{e}(j,t) \cdot \mathbf{w}(j,t) \qquad (1)$$

$$inh(j,t) = \mathbf{i}(j,t) \cdot \mathbf{w}'(j,t). \qquad (2)$$

The activation a of j at t is conceptualized as the probability of firing, defined as follows:

$$a(j,t) = \begin{cases} p(epsp,j,t) + \tau(j)p(epsp,j,t-1) \\ \qquad \cdot [1 - p(epsp,j,t)] \\ \quad - p(ipsp,j,t) \\ \quad \text{if } exc(j,t) \geq \epsilon(j,t) \\ \qquad \text{and } exc(j,t) > inh(j,t), \\ a(j,t-1) - k(j)a(j,t-1) \\ \qquad \cdot [1 - a(j,t-1)] \\ \quad \text{if } exc(j,t) < \epsilon(j,t) \\ \qquad \text{and } exc(j,t) > inh(j,t), \\ 0.0 \quad \text{if } exc(j,t) \leq inh(j,t) \end{cases} \qquad (3)$$

where $p(epsp,j,t) = L[exc(j,t)]$, $\tau(j)$ is a temporal-summation parameter $(0.0 < \tau(j) < 1.0)$, $p(ipsp,j,t) = L[inh(j,t)]$, $\epsilon(j,t)$ is a random excitatory threshold generated according to a Gaussian distribution with parameters μ and σ, and $k(j)$ is an activation-decay parameter $(0.0 < k(j) < 1.0)$. The term $p(epsp,j,t)$ is interpreted as the probability of occurrence of an excitatory postsynaptic potential (*epsp*) at j during t, whereas $p(ipsp,j,t)$ is interpreted as the probability of occurrence of an inhibitory postsynaptic potential (*ipsp*) at j during t. Function L is the logistic probability distribution with parameters γ and δ:

$$L(x) = 1/(1 + exp[(-x + \gamma)/\delta]). \qquad (4)$$

In the simulations, $\tau(j) = .1$, $k(j) = .05$, $\mu = 0.0$, $\sigma = 1.0$, $\gamma = .5$, and $\delta = .1$.

Learning Function

Let c: $N \times N \to \{0,1\}$ be the predicate "connected" (0 signifies "is not connected to" and 1 signifies " is connected to"), $E = \{i \in N \mid s(i) = 1\}$ and $I = \{i \in N \mid s(i) = 0\}$ be the sets of excitatory units and inhibitory units, respectively, where s: $N \to \{0,1\}$ is the function "releases" [0 signifies an inhibitory neurotransmitter (e.g., gamma-amino-butyric acid or GABA) and 1 signifies an excitatory neurotransmitter (e.g., glutamate)]. Let $\mathcal{E} = \{\langle i,j \rangle \in E \times N \mid c(i,j) = 1\}$ and $\mathcal{I} = \{\langle i,j \rangle \in I \times N \mid c(i,j) = 1\}$ be the sets of possible excitatory and inhibitory connections in a neural network, respectively. The mapping w: $\mathcal{E} \times T \to R$ is the learning function for excitatory presynaptic inputs, whereas w': $\mathcal{I} \times T \to R$ is the learning function for inhibitory presynaptic inputs. The networks trained in the simulations did not have inhibitory connections, so the learning function for inhibitory presynaptic inputs was not used.

The rules for implementing functions w and w' in the simulations are defined as follows. The connection weight w between a presynaptic element $i \in E$ and a postsynaptic element $j \in N$ at t is given by

$$w(i,j,t) = \begin{cases} w(i,j,t-1) \\ \quad + \alpha(j)a(j,t)d(t)p(i,t)r(j,t) \\ \quad \text{if } d(t) > 0, \\ w(i,j,t-1) \\ \quad - \beta(j)w(i,j,t-1)a(i,t)a(j,t) \\ \quad \text{if } d(t) \leq 0, \end{cases} \tag{5}$$

where $\alpha(j)$ is the acquisition rate for excitatory connections, $\beta(j)$ is the extinction rate for excitatory connection [in the simulations, $\alpha(j) = .5$, $\beta(j) = .035$, and $w(i,j,t = 0) = .01$], $d(t)$ is the reinforcing signal, $p(i,t) = [a(i,t) w(i,j,t-1)]/exc(j,t)$, and $r(j,t) = 1 -$

$sum[\mathbf{w}(j,t)]$. If j is a hidden NPE in the stimulus-selection subnetwork (i.e., a polysensory NPE), then $d(t) = \phi(t) + \omega(t) [1 - \omega(t)]$, where $\phi(t) = |avg[\mathbf{h}(t) - \mathbf{h}(t-1)]|$ and $\omega(t) = avg[\mathbf{v}(t) - \mathbf{v}(t-1)]$. In the last two equations, $\mathbf{h}$ is a vector of activations of all the hippocampal NPEs receiving inputs from the polysensory NPEs of the stimulus-selection subnetwork, whereas $\mathbf{v}$ is a vector of activations of all the VTA NPEs receiving inputs from the motor association NPEs of the response-selection subnetwork. On the other hand, if j is a motor NPE (i.e., an NPE in the response-selection subnetwork), then $d(t) = \omega(t)$. The use of the absolute value in function ϕ is based on the assumption that both the onset and offset of hippocampal NPEs influence sensory learning (i.e, the output of these NPEs is sensitive to sensory stimulus change). Because the networks trained in the simulations did not include any sensory NPEs, only $\omega(t)$ was used as the reinforcing signal.

The connection weight w' between a presynaptic element $i \in I$ and a postsynaptic element $j \in N$ at t is given by

$$w'(i,j,t) = \begin{cases} w'(i,j,t-1) \\ \quad + \alpha'(j)a(j,t)d(t)p(i,t)r(j,t) \\ \quad \text{if } d(t) > 0, \\ w'(i,j,t-1) \\ \quad - \beta'(j)w'(i,j,t-1)a(i,t)a(j,t) \\ \quad \text{if } d(t) \leq 0, \end{cases} \tag{6}$$

where $\alpha'(j)$ is the acquisition rate for inhibitory connections, $\beta'(j)$ is the extinction rate for inhibitory connections, $p(i,t) = [a(i,t) w'(i,j,t-1)]/inh(j,t)$, and $r(j,t) = 1 - sum[\mathbf{w}'(j,t)]$. The term $d(t)$ is defined as in the excitatory learning rule (see above).